（ 続紙 1 ）

京都大学	博士（情報学）	氏名	Mirzaei Maryam Sadat
論文題目	Partial and Synchronized Caption to Foster Second Language Listening based on Automatic Speech Recognition Clues （第二言語のリスニング訓練のための自動音声認識を用いた部分的かつ同期された字幕付与）		

（論文内容の要旨）

Central to the development of second language (L2) is the ability to perceive, process and comprehend the speech in the target language, which forms the bedrock of L2 listening skill. Listening is indeed a fundamental skill in L2 acquisition, which comes before speaking. It is the least explicit and essentially a transient and invisible process, hence the most sophisticated skill to master. To advance L2 listening skill, exposure to the authentic materials plays a crucial role. The advancement of ICT has promoted further opportunities for the application of contextualized and authentic materials, making them the mainstays of contemporary L2 learning education. Nevertheless, these materials, which are originally intended for native speakers of the target language, are often too difficult for L2 learners even at advanced levels. To facilitate the comprehension of such resources, captioning is widely used as an assistive tool for providing the text along with the speech. However, through the use of captions, learners tend to rely more on their reading skills, hence neglect the goal of training the listening skill.

This thesis attempts to solve this problem by introducing a novel captioning method, partial and synchronized captioning (PSC), as a tool for developing L2 listening skill. In this method, an ASR system is employed to align the words in precise timing with their respective speech signals in order to enable text-to-speech mapping and the caption is partialized by presenting words and phrases which are likely to hinder learner’s listening comprehension. Since there are various factors that lead to L2 listening difficulty, this study investigates the viability of using ASR errors as the predictor of difficulties in speech segments, thereby exploiting them to improve the baseline PSC system. Note that the human-annotated transcript is aligned with the ASR-generated transcript to realize synchronization and ASR error detection.

Chapter 1 provides an overview of the general topics of second language listening and the use of captioning to facilitate this process. It goes on to discuss the problems with the existing approaches and lays out the general motivation for and techniques used in the work presented in the following chapters.

Chapter 2 describes the overview of computer-assisted language learning (CALL) systems, as well as the use of ASR technology. It also focuses on the use of technologies in training L2 listening skill and discusses the different factors affecting L2 listening. This chapter provides a thorough introduction of different captioning methods and elaborates on the limitations of each method. Within this framework, this thesis presents the

importance of adopting partialization and synchronization for creating a new system of captioning.

Chapter 3 presents the idea of partial and synchronized caption as a tool that strives to mandate the shortcoming of previous methods. A trained ASR system allows for precise mapping between the text and the speech. With regards to partialization, the system relies on factors impeding the L2 listening process. In the baseline PSC system, partialization is performed based on well-known factors of speech rate, word frequency and specificity. This makes it straightforward to select difficult words and allows for caption adjustment through taking into account learners' vocabulary size and their tolerable rate of speech. Experiments demonstrate that the proposed method is able to realize comparable comprehension as the full caption while reducing the textual clues to less than 30%. The method is also able to address the requirement of L2 learners at different proficiency levels and can prepare them for listening in real-life situations.

Chapter 4 presents a comparative analysis of ASR errors and L2 learners' problems so that ASR errors can epitomize learners' listening difficulties with a particular audio. Given an erroneous output of the ASR system, we look for useful instances that can signal challenging speech segments for L2 listeners. The chapter reports on the analysis of ASR errors and the baseline PSC shown and hidden cases. This analysis provides hints for detection and inclusion of effective ASR errors into the PSC system, which is essential for achieving high accuracy in word selection to scaffold the learners. Annotation of ASR erroneous output led to the discovery of four effective categories of errors: homophones, minimal pairs, negatives, and breached boundaries to improve the choice of words in PSC. Experiments with L2 learners show that these categories are able to detect problematic speech segments and can be useful for enhancing the PSC system.

Chapter 5 extends the baseline PSC framework to encompass the features derived from the ASR errors and to enhance the word selection. In this view, two enhancements are made to improve the baseline system. The first improvement is removing easy cases by defining a secondary threshold for speech rate and word specificity features, taking into account the ASR correct or erroneous output, which allows for more effective pruning of the words coming to PSC. The second improvement is based on aggregating the four useful categories discovered in ASR errors, which lead to providing better choices of words to disambiguate the speech while listening. An experimental evaluation finds that the enhanced version of PSC is able to provide better clues for language learners while addressing most of their problems and being selected as a preferable caption.

Chapter 6 concludes the thesis with an overview of the baseline and the enhanced system of PSC and directions for future research.

注)論文内容の要旨と論文審査の結果の要旨は１頁を３８字×３６行で作成し、合わせて、３,０００字を標準とすること。
論文内容の要旨を英語で記入する場合は、４００～１,１００wordsで作成し
審査結果の要旨は日本語５００～２,０００字程度で作成すること。

（続紙 ２ ）

（論文審査の結果の要旨）

外国語（第二言語）のリスニング能力の訓練において字幕を提示すると、リスニングの支援にはなるが、学習者はしばしば字幕を読むことに注力し音声を聞かなくなるので、必ずしもリスニング能力の訓練にはつながらないという問題があった。本論文は、この問題に対して、自動音声認識技術を字幕テキスト作成でなく、（人でによる）字幕テキスト中の単語の選別と音声との同期に用いる新たな字幕提示法を提案するもので、主な成果は以下の通りである。

1. 従来の様々な字幕提示法の問題点を考察した上で、部分的かつ同期された字幕付与の方法を提案した。これは、学習者が聞き取りにくい単語のみを選別した上で、各単語を音声と同期し逐次的に提示するものである。これにより、音声と字幕の両方を補完的に活用したリスニングを実現する。聞き取りにくい単語の選別基準として、発話速度、単語の頻度及び専門性の３つの要因を用い、学習者のレベルに応じてしきい値を調整する。実験の結果、この提示法により、全部の単語を提示する場合と同様の理解を実現した上で、字幕が一切ない状況でのリスニングも高まることを確認した。

2. 自動音声認識システムの誤りを外国語学習者の聞き取り誤りと比較・分析した上で、音声認識誤りから外国語学習者の聞き取りが困難と予測される箇所を自動的に抽出する方法を考案し、その妥当性を確認した。

3. 上記2.に基づいた聞き取りにくい単語予測を組合せることで、部分的かつ同期された字幕付与を改善した。上記1.の３つの要因に基づいて単語を選択して提示する場合と比べて、学習者にとってより有用であることを確認した。

以上のように本論文は、自動音声認識システムの誤りが外国語学習者のモデルとなることを示した上で、外国語のリスニング能力訓練の高度化を実現する方法を提案したもので、学術上・実用上寄与するところが少なくない。よって、本論文は博士（情報学）の学位論文として価値あるものと認める。また、平成29年 2月17日に論文とそれに関連した内容に関する口頭試問を行った結果、合格と認めた。

注)論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。
更に、試問の結果の要旨（例えば「平成　年　月　日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」）を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。
要旨公開可能日：　　年　　月　　日以降